Does the Use of Mobile Devices (Tablets and Smartphones) Affect Stated Choice Behavior in Web Surveys? An Application to Environmental Valuation

Abstract:

Web surveys are becoming increasingly popular in survey research. Compared with face-to-face, telephone and mail surveys, web surveys may contain a different and new source of measurement error and bias: the type of device that respondents use to answer the survey questions. To the best of our knowledge, this is the first study that tests whether the use of mobile devices affects survey characteristics and stated preferences in a web-based choice experiment. The web survey was carried out in Germany with 3,400 respondents, of which 12 per cent used a mobile device (i.e. tablet or smartphone), and comprised a stated choice experiment on externalities of renewable energy production using wind, solar and biomass. Our main finding is that survey characteristics such as interview length and acquiescence tendency are affected by the device used. In contrast to what might be expected, we find that, compared with respondents using desktop computers and laptops, mobile device users spent more time to answer the survey and are less likely to be prone to acquiescence bias. In the choice experiment, mobile device users tended to be more consistent in their stated choices, and there are differences in willingness to pay between both subsamples.

Keywords:

survey format; stated choice experiment; mobile device; renewable energy; acquiescence bias; survey quality

1 Introduction

Stated preference surveys are increasingly being conducted online, which can be attributed to increased internet penetration rates and the advantages online survey formats offer over alternative survey formats (Dillman, et al., 2009, Manfreda and Vehovar 2008). The online format allows for surveys to be administered to large samples in a short period of time at a relatively low cost. Online studies permit efficient and novel ways to convey information regarding the valuation context, for example using multi-media tools, and to efficiently control the survey flow. Online formats also enable researchers to easily collect additional information on response conditions and behavior (paradata) such as response times and latencies, which may be used to explain variation in choice behavior (Campbell et al. 2012, 2013; Dellaert et al. 2012). Provided that the penetration of the internet and the availability of internet-based services will continue to increase, it is conceivable that web surveys will become the dominant survey format of the future.

Therefore, there is an interest in understanding how online stated preference surveys compare to other survey formats in terms of representativeness and response behavior. Findings thus far are mixed. Compared to alternative survey formats, in terms of representativeness, web surveys may produce samples that are unbalanced towards male respondents, that are younger, more highly educated and have higher income (Kwak and Radler, 2002, Lindhjem and Navrud, 2011, Marta-Pedroso et al., 2007, Olsen, 2009). However, differences are study specific. In terms of response behavior, Lindhjem and Navrud (2011), Nielsen (2011) and Marta-Pedroso, et al. (2007) find no significant differences in mean willingness to pay (WTP) in comparisons of online and face-to-face surveys applying the contingent valuation method. In a comparison of online and mail survey formats using choice experiments, both Olsen (2009) and Windle and Rolfe (2011) could not reject the hypothesis of equal WTP estimates. However, after controlling for sample frame and self-selection effects, Morrison et al. (2013) recently found that the online survey resulted in WTP estimates that were, on average, 30% lower than those derived via a mail survey.

This study differs from previous comparative studies of survey formats. Instead, we focus entirely on respondents to a web survey, and investigate whether the device used for completion has an impact on response behavior. In particular, this is, to the best of our knowledge, the first time that the impact of the use of mobile devices (mobile phones, tablets) is compared to using desktop computers and laptops (desktop/laptop users) in completing a stated preference survey. The recent years have seen a rapid expansion of the use of internetenabled mobile devices such as smartphones and tablets. If the internet is increasingly accessed via such devices, it can be expected that online surveys will also be increasingly be completed on smartphones and tablets. Research on the impacts of using mobile devices to complete surveys is still in its infancy (see, e.g., Callegaro 2010, Peytchev and Hill 2010, Buskirk and Andrus 2012, Millar and Dillman 2012 for notable exceptions). Peytchev and Hill (2010), for example, do not find differences regarding cognitive processing and use of pictures comparing mobile web surveys and other survey modes. However, they find that users of mobile phones are less likely to provide text input and show differences in response behavior, if the survey questions extend beyond the screen. However, this study is limited by a small sample size of 92 respondents. Millar and Dillman (2012) conducted an experimental study with 600 undergraduate students, in which they tested whether the response rate of smartphone users increases, if respondents are explicitly encouraged to use the smartphone for answering the online questionnaire. This treatment group was compared to respondents, who were requested to take part in an online survey, and a third group that could choose between answering an online questionnaire or a paper copy of it. Millar and Dillman (2012) do not find that explicitly requesting to use the smartphone has an effect on the response rate.

The impact of using mobile devices on response behavior in stated choice experiments is difficult to predict, because it may depend on a large array of unobserved factors. For example, one may surmise that the use of mobile devices implies completing the survey while being mobile, for example, during the daily commute to work. We would then expect that the survey will be interrupted more frequently and that respondents are more distracted, resulting in a greater error variance compared to using desktop computers and laptop. However, desktop/laptop users may equally be distracted. In the case of respondents using laptops, the circumstances may be similar, for example if laptops are used on a train or in a cafe. Regarding the use of stationary desktop computers, the use of different programs and email that are competing for their attention, or the radio or TV show playing in parallel, may be examples of potential sources of distraction that could impact on the accuracy of choices made.

An observable difference between desktops/laptops and different types of mobile devices is related to screen size. Tablets and particularly smartphones typically have a smaller screen size, which may require respondents to either zoom in and out of choice cards frequently, or to scroll laterally to compare attribute information between different alternatives. Again, giving the apparent difficulty in accessing the whole information entailed in a choice task, one may conjecture that smaller screen size is associated with greater error variance. However, the difficulty in accessing the information on a mobile device may equally prompt respondents to expend more effort on taking in the information, and on making the decision, which may result in reduced error variance. Differences in WTP estimates may arise if respondents employ different decision rules, or the same rules to a differing degree. For example, it may be conceivable that non-attendance to attributes differs between users and non-users of mobile devices. Similarly to error variance, we are not able to form any directional hypotheses regarding differences in preferences and estimates of WTP.

Against this backdrop, this study is largely exploratory in the sense that we test for differences in various observable survey characteristics such as interview length and acquiescence tendency as well as stated choice behavior between users and non-users of mobile devices. The data was obtained in a web survey on renewable energy production in Germany, which included a stated choice experiment on externalities of the renewable energy production from wind, solar and biomass energy sources. Twelve per cent of the 3,400 respondents used a mobile device (tablet or smartphone) to answer the survey. We use a pairwise matching approach to make the subsamples of users and non-users of mobile devices comparable. In short and contrary to what might be expected, our findings indicate that survey quality and choice consistency in the choice experiment tends to be higher for users of mobile devices compared with non-users. In the following, we first describe the study's background, the stated choice experiment, data collection and data. We proceed with presenting results regarding survey characteristics and stated preferences. The paper concludes with a discussion of our approach and findings.

2 Study and Data

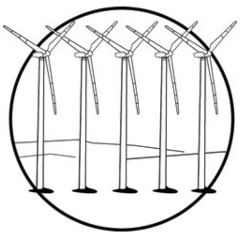
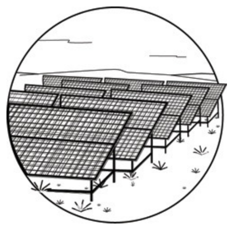
In the following, we first present background information on our survey and design of the stated choice experiment. Second, we give an overview on the data and pairwise matched sample used to compare desktop/laptop users and mobile device users.

2.1 Study design and survey administration

Apart from the stated choice experiment on renewable energy expansion in Germany, the questionnaire comprises questions concerning respondents' exposure to renewables, attitudes, acceptance and fairness aspects regarding the expansion of renewable energies in Germany, and socio-demographics. Three renewable energy sources were considered: wind energy, solar energy, and biomass. At the beginning of the survey, respondents were shown pictograms and definitions of these renewables (see Table 1). It was also clarified that the survey focused on renewables in the open landscape and did not consider, for example, energy production in urban areas using solar panels.

Six focus groups in different towns spread over Germany were conducted in October 2012 to assess understanding and acceptance of the questionnaire. After discussing perceived advantages and disadvantages of renewable energies in Germany, participants completed an earlier version of the choice experiment and subsequently reported their views and impressions. Based on these comments, the choice experiment was revised. In particular, the number of choice sets and attributes was reduced. The revised questionnaire was tested in two pilot studies. The first study (N=74) was conducted with colleagues and the second (N=100) with members of the access panel provided by the survey organization that also carried out the main survey.

Table 1: Definition of renewable energy sources used in the survey

		
Wind energy refers to electricity generation with single wind turbines and wind farms exclusively onshore.	**Solar energy** refers exclusively to the production of electricity with photovoltaic systems in the open landscape, so-called solar fields.	**Biomass** refers to the production of biogas and its electricity and includes both the biogas plant as well as the cultivation of the required biomass (such as corn).

In the choice experiment respondents had a choice between four labelled alternatives. Three alternatives described future options for renewables expansion of wind energy, solar energy or biomass within a 10-kilometre surrounding of their place of residence. The labelled alternatives were introduced using the pictograms and definitions shown in Table 1. In addition, respondents could choose a future status quo alternative with zero cost to them. This alternative, which was labelled “no influence on renewable type”, detailed the most likely future renewables expansion scenario in the absence of any further policy intervention. The choice attributes are reported in Table 2. They relate to the minimum distance of the production sites to the edge of town, the size and number of the production sites, whether new high-voltage transmission lines are build overhead or underground, and the landscape area set aside for landscape protection. The price attribute was a surcharge or rebate to the energy bill.

Figure 1gives an example of a choice set. Respondents were requested to choose in each choice set their preferred alternative regarding the renewable energy future within a 10-kilometre radius of their place of residence (highlighted green in the second last row of the choice set), and their least preferred alternative (last row of the choice set, highlighted red). As the choice refers to changes within a 10- kilometre radius from the place of residence, respondents living in big cities were asked to instead think about the landscape surrounding the city assuming that they might use the landscape for recreational purposes. The price attribute was given not only in terms of price changes per month, but also in terms of the annual change of the energy bill.

Table 2: Attributes and attribute levels

Attribute	**Alternative**	**Level**
Minimum distance to residential areas		300 / 600 / **900** / 1600 / 2500
Size of production site	Wind	small (5-10 turbines) / **medium** (18-25 turbines), large (35-50 turbines)
	Solar	small (1-10 football fields) / **medium** (20-60 football fields), large (100-150 football fields)
	Biomass	small (1-3 fermentation tanks) / **medium** (5-8 fermentation tanks), large (15-25 fermentation tanks)
Number of production sites		1 / 2 / **3** / 4 / 5
Share of landscape not used for production (in %)		10 / 20 / **30** /40 / 50
Long-distance Transmission lines		**overhead** / underground
Cost in Euro (surcharge or rebate to energy bill)		-10(-120) / -5(-60) / +2(24) / +7(84) / +14(168) / +23(276)

Note: Levels of the future status quo alternative are written in bold.

In order to combine the attribute levels into choice sets, we generated a Bayesian efficient design with labelled alternatives using Ngene software. As the optimization criterion we used the C-error, which allows minimisation of the variance of the sum of the marginal WTP estimates (Scarpa and Rose 2008). The prior values were taken from model estimates based on data collected in the focus-groups and the pilot studies. The resulting design had 24 choice sets that were blocked into four blocks with each 6 choice sets. The order of appearance of choice sets was randomised. Additionally, the order of the first three non-status quo alternatives was randomised across respondents, that is, the order of alternatives was the same for each respondent but differed across respondents.

In the present case, our survey was optimized for the use with mobile devices. Web surveys can be optimized for mobile devices: generally, this means that for example lists with response options are dynamically adjusted to the size of a mobile device. This allows users of mobile devices to more easily navigate through the survey. However, optimization has its limits concerning the display of larger survey components such as choice sets. While a choice set may be displayed in full on a mobile device screen, it may not be readable due to small screen sizes. This is more likely if the choice set is larger as in our case of sets with four alternatives and six attributes (see Figure 1). Therefore, some respondents using mobile devices will probably have had to zoom and move the choice sets to access all the information contained in the sets and also to tick the alternative they prefer.

	Electricity from wind	**Electricity from solar**	**Electricity from biomass**	**No influence on renewable type**
Minimum distance to town	600m	2500m	300m	900m
Size of production sites	large (35-50 turbines)	large (15-25 fermentation tanks)	small (1-10 football fields)	medium
Number of production sites	4	5	5	3
Protection of landscape	20%	50%	10%	30%
Transmission lines	Underground	Underground	overhead	overhead
Change in energy bill	+14€ (+168€)	-5€ (-60€)	+14 € (+168 €)	0 €
I choose				
…. best option	☐	☐	☐	☐
…. worst option	☐	☐	☐	☐

Figure 1: Example of choice set

The data resulted from a nationwide online survey that took place in September and October 2013. Participants were members of an access panel. A shopping voucher for an online mail order company amounting to 3.50 Euro was used as an incentive to complete the interview. In total 12,833 panel participants were invited to take part. Of these, 220 could not be admitted to the survey as quota restrictions were already fulfilled (a quota system for age and sex was applied), 4,027 persons took part in the survey, and 3,400 completed the questionnaire. After inspection of the data, 3,396 usable interviews remained. According to the AAPOR-Standard Definitions (standard RR1, see AAPOR 2009) this corresponds to a response rate of 27.9%.

2.2 Original Sample and Pairwise Matched Sample

Table 3 gives an overview of the subsamples of respondents (desktop/laptop users and users of mobile devices). Mobile-device users are further differentiated into tablet users and smartphone users. In our study, 11.6% of the respondents used a mobile device to answer the web survey (389 respondents of overall 3,344 respondents), of which 6.1% (N=203) used a tablet and 5.5% (N=185) a smartphone. The type of tablets and smartphones most often used across the sample were iPad (36.5%) and iPhone (20.8%) as well as various Samsung phone models (15.7%) and Samsung tablet models (7.5%). Comparing the first and third column in Table 3, it can be seen that mobile device users are more likely to be female, younger and to have a lower education compared to those respondents, who completed the survey on PCs/laptops. It can further be seen that the differences are more pronounced for the subsample of smartphone users (last columns). In this subsample the share of females is considerably higher than in the subsamples of desktop/laptop users and tablet users, mean age is lower by at least 10 years, and also the share of respondents with higher education is remarkably lower.

Table 3: Overview on original and pairwise matched sample

	Whole sample	Pairwise matched sample			
	Desktop computer/laptop users	Desktop computer/laptop users	Mobile-device users		
			All	Tablet	Smartphone
	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max
Gender (1=women)	.45 (.50) 0/1	.50 (.50) 0/1	.50 (.50) 0/1	.48 (.50) 0/1	.51 (.50) 0/1
Age in years	43.55 (14.18) 18/84	37.37 (12.26) 18/81	36.75 (12.23) 17/78	41.52 (12.66) 18/78	31.50 (9.30) 17/69
Education (1=upper secondary+)	.70 (.46) 0/1	.67 (.47) 0/1	.67 (.47) 0/1	.69 (.46) 0/1	.64 (.48) 0/1
N	2955	389	389	203	185

A probit model for use of a mobile device, which is presented in Table 4, shows that the differences between the subsamples “desktop/laptop users” (first column) and “mobile-device users” (third column) are statistically significant with respect to respondents’ age and education. The difference for gender is not statistically significant (but it is statistically significant in a logit model based on the same variables presented in Table 4).

Table 4: Probit model for use of a mobile device

	Mobile device (1=yes, 0=no)
Gender (1=women)	0.09 (1.52)
Age in years	-0.02*** (-8.94)
Education (1=upper secondary+)	-0.14* (-2.22)
Constant	-0.32** (-2.85)
LL	-1,156.330
McFadden R^2	0.038
N	3,344

Note: z-values in parentheses; *** p<0.001, ** p<0.01, * p<0.05, $^{+}$ p<0.10.

We used a pairwise matching approach to make the subsamples comparable regarding the respondent characteristics gender, age, and education. For age we constructed a categorical variable with four categories representing 25% intervals: “17 to 30 years”, “31 to 43 years”, “44 to 53 years”, and “older than 53 years”. We then considered each possible combination of the two binary variables gender and education and the categorical variable age (“female, lower education, 17 to 30 years”, “male, lower education, 17 to 30 years”, etc.). For each combination we drew a random sample from the subsample of desktop/laptop users, which matches the size of the corresponding group within the subsample of mobile device users (e.g., random sample of 43% of all respondents without mobile device who are female, have lower education and are between 17 and 30 years old). The final pairwise matched sample of 389 respondents of desktop computer/laptop users is shown in the second column in Table 3. The slight difference between the pairwise matched samples (second and third column) regarding age are due to the fact that the variable age is presented in years, whereas a categorical variable was used for the pairwise matching.

3 Econometric Approach

All choice models are estimated in WTP space following Train and Weeks (2005) and Scarpa et al. (2008). The modelling approach is based on the random utility theory, with a utility function U for respondent n and alternative i in choice task t characterised by price p and non-price attributes x of the experimental design, and a random error term ε:

$$U_{nit} = -\alpha_n{}' p_{nit} + \beta_n{}' x_{nit} + \varepsilon_{nit} \tag{1}$$

Where α' and β' are parameters to be estimated and ε is assumed to be identically and independently distributed (*iid*) and related to the choice probability with a Gumbel distribution with respondent specific error variance $\mathrm{Var}(\varepsilon_{ni}) = \mu_n^2(\pi^2/6)$, with μ_n being a respondent specific scale factor.

Train and Weeks (2005) show that equation (1) can be divided by μ_n to derive a scale-free and behaviorally equivalent utility function with a new error term that is constant across individuals:

$$U_{nit} = -(\alpha_n/\mu_n)' p_{nit} + (\beta_n/\mu_n)' x_{nit} + \varepsilon_{nit} \tag{2}$$

Where ε_{nit} is *iid* with constant error variance $\pi^2/6$. Substituting $\gamma_n = \alpha_n/\mu_n$ and $c_n = \beta_n/\mu_n$ in equation (2) as the parameters to be estimated provides what Train and Weeks (2005) call the model in preference space. Exploiting the fact that WTP is $w_n = c_n/\gamma_n$, the utility function in WTP space can be written as:

$$U_{nit} = -\gamma_n{}' p_{nit} + (\gamma_n w_n)' x_{nit} + \varepsilon_{nit}. \tag{3}$$

Denote the sequence of choices over T_n choice tasks for respondent n as y_n, i.e. $y_n = \langle i_{n1}, i_{n2}, \ldots, i_{nT_n} \rangle$. In a random parameter logit (RPL) model, heterogeneity across respondents is introduced by allowing γ_n and w_n to deviate from the population means following a random distribution. Conveniently, as opposed to models in preference space, the distribution of WTP can be estimated directly. In a RPL model, the unconditional or mixed logit choice probability of respondent n's sequence of choices is the integral of the logit formula over all possible values of γ_n and w_n:

$$\Pr(y_n|\gamma_n, w_n) = \int \prod_{t_1=1}^{T_n} \frac{\exp(-\gamma_n{}' p_{nit} + (\gamma_n w_n)' x_{nit})}{\sum_{j=1}^{J} \exp(-\gamma_n{}' p_{njt} + (\gamma_n w_n)' x_{njt})} f(\eta_i|\Omega) d\eta_i \tag{4}$$

where $f(n_i|\Omega)$ is the joint density of parameter vector for price and K non-price attributes [γ_n, w_{n1}, w_{n2}, … , w_{nK}], η_i is the vector comprised of the random parameters and Ω denotes the parameters of these distributions (e.g. the mean and variance). This integral does not have a closed form and thus requires approximation through simulation (Train, 2003), in our case using 1,000 Halton draws. The price attribute parameter is assumed to follow a lognormal distribution, the WTP parameters are assumed to follow a normal distribution.

The variance of the error term may differ between subgroups of respondents, in our case between users of mobile devices and respondents who used desktop/laptops to complete the survey. Relative differences in error variance can be identified by allowing scale to differ between subgroups, such that the deterministic part of equation (3) becomes

$$V_{nit} = (\lambda_{desk/laptop} + \lambda_{mob} I_{mob}) \left(-\gamma_n{}' p_{nit} + (\gamma_n w_n)' x_{nit}\right) \tag{5}$$

Where $\lambda_{desk/laptop}$ and λ_{mob} are *relative* scale parameters for desktop/laptops and mobile device users that are inversely related to error variance, I_{mob} is an indicator variable taking one if individual n used mobile devices, else zero. $\lambda_{desk/laptop}$ is set to one to allow identification.

The error variance may also differ between individuals of a subgroup as a function of respondent specific characteristics, in our case screen size S_n. In this case, we specify a heteroscedastic logit model (Swait and Adamowicz, 2001, DeShazo and Fermo, 2002) using

$$V_{nit} = \lambda_n(S_n) \left(-\gamma_n{}' p_{nit} + (\gamma_n w_n)' x_{nit}\right) \tag{6}$$

where $\lambda_n(S_n) = \exp(\theta_1 S_n + \theta_2 S_n^2)$, that is, we assume a quadratic relationship between screen size and scale (error variance). This mirrors an expectation that respondents who use small screens are less consistent in their choices up to a threshold of screen size, after which error variance increases, possibly because respondents require less effort in accessing the relevant information that characterises the alternatives. The exponential function ensures positive scale and has excellent convergence properties (DeShazo and Fermo, 2002). In the is, S_n enters as a zero-centred variable, implying that at the sample mean $\lambda_n(S_n) = 1$.

4 Results

We first present results comparing desktop/laptop users and mobile device users with respect to several survey characteristics. Some of these characteristics such as acquiescence tendency represent indicators of survey quality. We subsequently proceed with the results of the choice experiment investigating differences in stated preferences between subsamples.

4.1 Group Comparison Regarding Survey Characteristics

Table 5 contains for each subsample the descriptive statistics for several survey characteristics and response patterns. In order to find out, whether differences between subsamples are statistically significant, results of Mann-Whitney U tests and Chi2-test are presented. It can be seen that, unsurprisingly, smartphone users have a remarkably lower screen size than tablet users. Respondents' smartphones have a mean screen size of 10 cm and tablets have a mean screen size of 25 cm, which is closer to the screen size of a standard PC and laptop. However, differences between respondents using desktop/laptops and those using a mobile device are more pronounced for smartphone users compared with tablet users. This is the case for all variables in Table 5, except interview time.

Mobile device users tend to complete the survey somewhat later in the day. However, the difference is only statistically significant for tablet users. A more detailed analysis reveals that most respondents, irrespective of using a desktop/laptop or mobile device, answer the survey at the evening (between 6pm and 11pm, 46.4%) or afternoon (between 12am and 5pm, 34.7%), followed by morning (between 6am and 11am, 16.2%) and night (between 12pm and 5am, 2.7%). Mobile device users are more likely to interrupt the survey (4% versus 8 %).The difference is statistically significant for smartphone users (4% versus 10%) only (not for tablet users, 4% versus 6%). Further, mobile phone users spend, on average, more time to answer the survey. The difference in mean interview length amounts to 11 minutes (33 minutes versus 44 minutes) and is statistically significant for both tablet (47 minutes) and smartphone users (42 minutes). In order to account for outliers and the large variance, Table 5 also reports mean interview length without the lowest and highest 5% in each subsample. While the difference in interview length between desktop/laptop and mobile device users decreases to 4 minutes, all differences are still statistically significant. The higher values for interview length might indicate a higher quality of responses, possibly because respondents read the questions and choice tasks more carefully. On the other hand, it might show that it is more difficult to answer the survey questions for mobile device users due to, for example, smaller screen size. This is, however, less likely to apply in our study, because respondents were recruited from an access panel, which implies that they are experienced in answering web surveys.

We are generally aware that it is by no means easy to define and investigate the quality of survey responses (Lyberg and Biemer 2007). Nonetheless, we calculated the respondents' acquiescence tendency as an indicator of response quality. The tendency to agree in a survey regardless of the content of the survey question is a well-known bias in survey research

(Schaeffer and Presser 2003). It might have several causes including differences between respondents regarding cognitive skills; in our study differences in the acquiescence tendency between desktop/laptop and mobile-device users might be also interpreted as differences in respondents' effort to answer the survey question, that is, extreme response patterns such as always agreeing or disagreeing are more/less likely.

Table 5: Overview on survey characteristics per subsample

	Whole sample	Pairwise matched sample			
	Desktop computer/laptop users	Desktop computer/laptop users	Mobile device users		
			All	Tablet	Phone
	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max
Screen size in cm			17.47 (7.43) 7.1/25.7 N=374	24.51*** (1.49) 17.8/25.7 N=193	9.97*** (1.44) 7.1/13.5 N=180
Interview time (in hours)	15.58** (4.52) 1/23 N=2955	15.63*/** (4.52) 1/23 N=389	16.07* (5.10) 1/23 N=389	16.49** (4.80) 1/23 N=203	15.59 (5.40) 1/23 N=185
Interview interrupt-ted (1=yes)	.05* 0/1 N=2955	.04* 0/1 N=389	.08* 0/1 N=389	.06 0/1 N=203	.10* 0/1 N=185
Interview length in minutes	31.91***(75.69) 4.48/ 2815.98 N=2955	33.38*** (102.42) 6.02/1739.15 N=389	44.39*** (102.17) 7.28/1390.03 N=389	46.69*** (130.69) 7.28/ 1390.03 N=203	41.98*** (57.01) 10.60/514.42 N=185
Interview length in minutes (without lowest/highest 5%)	26.78***(9.46) 14.03/59.98 N=2598	25.37*** (9.01) 14.03/59.48 N=349	28.74*** (9.25) 14.03/55.80 N=353	27.91*** (8.91) 15.35/55.80 N=187	29.70*** (9.58) 14.03/55.32 N=165
Acquiescence tendency	.60* (.21) 0/1 N=2952	.60* (.22) 0/1 N=389	.58^{+} (.21) 0/1 N=388	.60 (.21) 0/1 N=202	.55* (.21) 0/1 N=185
Share of status quo choices in %	10.70	10.41	10.24	9.93	10.54

Note: *** p<0.001, ** p<0.01, * p<0.05, $^{+}$ p<0.10. Significance levels for group comparison between the subsample without mobile and the subsamples with mobile are based on a Mann-Whitney U Test. Significance tests for the variable interview interrupted are based on a Chi2-Test. In order to calculate the acquiescence tendency, for each respondent we summed up the agreement answers (1=agree/completely agree) to eight questions with a four-point agreement scale and divided this sum by the number of items. It follows that a value of 0 means that a respondent has never agreed (agree or completely agree) and a value of 1 that she/he has always agreed.

We see in Table 5 that mobile device users have a lower acquiescence tendency than non-users (0.58 versus 0.60), but the difference is only statistically significant for smartphone users (0.55 versus 0.60); there is no significant difference between desktop/laptop users and tablet users (both with a value of 0.60). Note that the negative effect of mobile phones on acquiescence tendency is also statistically significant at the 5% level if we control for respondents' gender, age and education, based on an ordinary least square regression with acquiescence as dependent variable and gender, age, education as well as use of smartphone as independent variables (results are available from the authors).

With regard to the stated choice experiment, the tendency to choose the status quo or zero price alternative in a forced-choice design can be interpreted as an opt out response, among others (Kontoleon and Yabe 2003). Our data do not show significant differences in the share of future status quo choices between desktop/laptop users and smartphone users. The share is around 10% in each subsample.

4.2 Stated Choices Taking Device into Account

Table 6 shows the results of the stated-choice models. All models are highly statistically significant. All attribute coefficients carry the expected sign, and the alternative specific constants and most attribute coefficients are significant at the 10% level or greater. Exceptions are renewable expansion via large areas (*Area_l*) and number of sites (*#sites*) in the mobile device subsample. Across all models, there is a tendency to move away from the future status quo. For reasons not explained by attributes, respondents of the desktop computer/laptop and mobile device subsamples prefer renewable energy expansion in their area using solar, wind and biogas over the future status quo. Relative to medium sized areas assigned to renewable energy expansion, larger areas are associated with a disutility, while smaller areas are associated with a utility gain. A greater distance of sites dedicated for renewables, and sites being connected to the grid underground rather than above ground are preferred. Respondents also prefer to see renewables being spread to a greater number of sites. Finally, respondents have preferences for larger areas specifically assigned to landscape conservation. The standard deviation parameters are significant for several of the attributes, indicating the presence of significant heterogeneity in WTP for most attributes, as well as for changes in the energy bill (*Price*).

By visual inspection only, it is apparent that WTP related to the type of technology on which energy expansion focuses (i.e. WTP for the ASCs) is broadly similar for desktop/laptop and mobile device users, and that some of the estimated mean WTP values for attributes differ between subsamples. A Poe et al. (2005) test confirms significant differences in mean WTP for two of the attributes (expansion via large production areas (*Area_l*), and having underground transmission lines (*Grid*)).

Table 6: Choice Model results

	Model 1				**Model 2**				**Model 3**				**Model 4**			
	RPL WTP space desktop/laptop users				RPL WTP space mobile device users				RPL WTP space whole sample two scale groups				RPL WTP space whole sample het. scale as a function of screen size			
	Mean		StD		Mean		StD		Mean		StD		Mean		StD	
ASCB	8.71	***	-		7.46	**	-		7.7	***	-		10.1	***	-	
	(2.75)				(2.67)				(1.84)				(3.14)			
ASCS	25.9	***	-		24.8	***	-		25.0	***	-		27.7	***	-	
	(3.05)				(2.78)				(1.97)				(3.26)			
ASCW	20.1	***	-		15.1	***	-		16.7	***	-		18.3	***	-	
	(2.91)				(2.65)				(1.86)				(3.31)			
Area_l	**-0.51**	***	0.66		**-0.22**		0.92	***	-0.35	***	0.74	***	-0.22		0.99	***
	(0.16)		(0.42)		(0.13)		(0.23)		(0.10)		(0.21)		(0.15)		(0.23)	
Area_s	0.61	***	1.49	***	0.63	***	1.11	***	0.58	**	1.36	***	0.66	***	0.71	
	(0.17)		(0.23)		(0.14)		(0.23)		(0.10)		(0.15)		(0.13)		(0.47)	
Distance	0.43	***	0.40	**	0.35	***	0.29	**	0.39	***	0.36	***	0.35	***	0.13	
	(0.08)		(0.17)		(0.06)		(0.12)		(0.05)		(0.1)		(0.06)		(0.25)	
Grid	**0.81**	***	0.95	***	**0.528**	***	0.64	***	0.66	***	0.74	***	0.48	***	0.52	***
	(0.14)		(0.21)		(0.11)		(0.18)		(0.09)		(0.14)		(0.11)		(0.18)	
Landscape	1.64	***	5.53	***	0.97	**	2.85	***	1.26	***	4.3	***	0.97	***	1	
	(0.52)		(0.69)		(0.36)		(0.64)		(0.31)		(0.41)		(0.3)		(0.93)	
Site #	0.92	*	0.99		0.56		1.03		0.67	*	0.02		0.72		1.94	**
	(0.54)		(1.64)		(0.46)		(1.23)		(0.34)		(1.2)		(0.50)		(0.93)	
Price	-3.07	***	0.58	***	-2.88	***	0.82	***	-3.03	***	0.69	***	-2.65	***	0.94	***
	(0.08)		(0.11)		(0.09)		(0.12)		(0.07)		(0.08)		(0.23)		(0.13)	
$\lambda_{desk/laptop}$									1							
									(fixed)							
$\lambda_{mobile\ dev}$									1.15							
									(0.1)							
θ_1															0.01	
															(0.1)	
θ_2															-0.61	
															(0.40)	
# of obs	2334				2334				4668				2334			
Rho2	0.15				0.18				0.16				0.18			
LogL	-2736.5				-2658.7				-2736.3				-2657.4			

Note: *,**,*** significant at 10%,5%,1% level; in bold: differences in mean WTP between desktop/laptop and mobile device users significant at 10% level based on Poe et al. (2005) test. The attribute levels were scaled before entering the analysis. Parameters for size of production sites (*Area_l* and *Area_s*) and for transmission line type (*Grid*) must be multiplied by 10 to obtain WTP values for changes from the status quo. The parameter for minimum distance to town (Distance) reflects WTP per 100 meters, the parameter for area set aside for landscape protection (*Landscape*) reflects WTP per 10%.

Model 3 is a pooled model of desktop/laptop and mobile device subsamples, allowing for differences in scale between respondents who completed the survey on a desktop/laptop and those who used smartphones or tablets. The scale parameters associated with smartphone and tablet users are larger in magnitude, suggesting that on average respondents using mobile devices show greater consistency in their choices. However, differences in scale between desktop/laptop and smartphone users are not statistically significant ($p = 0.11$; $t(1) = 1.6$).

While error variance is therefore not found to significantly differ on average *between* users and non-users of mobile devices, there may be differences *within* the mobile device subsample that are related to screen size. Model 4 represents no statistical improvement in model fit over Model 2, and both θ_1 and θ_2 are insignificant. There may not be sufficient variation in our relatively small sample; nevertheless, for illustrative purposes we show the implied negative quadratic relationship between screen size and error variance in Figure 2. It suggests that there may be a threshold when choice consistency is negatively affected by screen size, but that there are few differences once a mobile device has a certain screen size. The mean screen size is 24.5 cm for tablets and 10 cm for smartphones. Comparing error variance at the mean level of screen size for tablets and smartphones, it can be easily seen that differences are small.

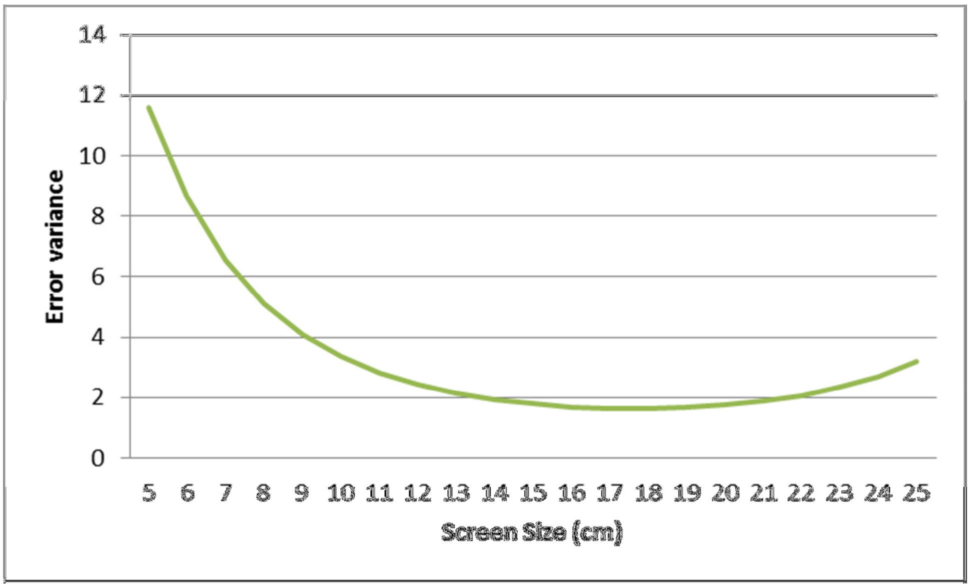


Figure 2: Estimated relationship between screen size and error variance

5 Discussion and Conclusions

To the best of our knowledge, this is the first study that tests whether the use of mobile devices affect survey characteristics and stated preferences in a web-based choice experiment study in an environmental valuation context. Our main finding is that survey characteristics such as interview length and acquiescence tendency as well as choice consistency in the stated choice experiment are affected by the device used; yet the differences for choice consistency are not statistically significant. In contrast to what might be expected, we find that, compared with respondents using a desktop/laptop, mobile device users spent more time to answer the survey and are less likely to be prone to acquiescence. Overall, these differences are more pronounced for mobile phone users compared with tablet users. Error variance was found to be lower for the mobile device subsample indicating greater choice consistency among mobile device users, albeit the difference was not statistically significant. WTP generally tends to be lower for mobile device users, and is significantly different for two of the choice attributes between subsamples.

Mobile devices might be associated with answering the survey “on the way” and difficulties in viewing and accessing the information in the questionnaire, especially with the smaller screen size found in smartphones. This might be considered as disadvantages of mobile devices compared with the use of desktop computers or laptops. It has to be stressed that any web survey has to be optimized for mobile device in order to guarantee similar visual experiences for users with and without mobile device (see, e.g., Burskirk and Andrus 2012 for approaches how to implement smartphone survey). In the present survey, the questionnaire has been optimized for mobile devices, and this might be one reason why we do not find indications of lower survey quality. However, the usefulness of the optimization for smartphones may be limited for the choice-experiment part of the survey, because the displayed choice sets may have been too small to comprehend without zooming and/or scrolling. Nevertheless, our results even suggest higher survey quality for mobile device users in general and smartphone users in particular shown by lower acquiescence tendency and higher choice consistency. With respect to choice consistency, however, we also find that consistency is somewhat lower for users of very small smartphones (curvilinear relationship between screen size and error variance).

This study has some limitations. First, the pairwise matching we used is a quite simple approach. Although our main findings seem to be robust if we repeat the procedure (i.e. draw another random sample), we aim to compare our results with those based on more complex sampling approaches such as propensity score matching (Rosenbaum and Rubin 1983, Morgan and Winship 2007). Second, the number of characteristics, which we have considered for matching (i.e. gender, age, and education), are limited, too. There might be further characteristics that affect the use of a mobile device to answer web surveys as well as the stated choices. This might be one reason why we find differences in stated preferences and WTP between desktop/laptop and smartphone users. These differences are difficult to explain and, ideally, a sampling approach would include all relevant characteristics to ensure that we compare subsamples that have similar preferences for the environmental good at hand. Another possible explanation for differences in preferences and WTP is the possibility that respondents to both samples applied different decision rules and information processing strategies. For example, non-attendance to choice attributes might differ between desktop/laptop and mobile device users. This should be followed up in future studies. Third, given the large-scale nature of our survey (N=3,344) we have a large number of respondents using a mobile device in the data set (N=389 or 12% of all respondents). However, we find strong variation and differences within the subsample of mobile device users (i.e. tablet users versus smartphone users). Therefore, an even larger sample of mobile device users is desirable to investigate the heterogeneity among mobile device users regarding the impact of screen size on error variance, for instance. At this point our results are only indicative. Fourth, in future studies investigating effects of mobile devices on responses in web surveys, the type of device should be taken into account in the sampling process. This will solve most of the problems mentioned above. Further, respondents in our study were members of an access panel and, hence, they are experienced with answering web surveys. Differences between desktop/laptop and mobile-device users might be larger if “inexperienced” respondents answer the survey.

Notwithstanding limitations, this study is a first step analyzing effects of using mobile devices in web surveys on environmental valuation and find interesting differences between respondents using a desktop computer/laptop and mobile device. This opens the way for more detailed studies on the use of mobile devices in web surveys. Our study also adds evidence to the literature that demonstrates the usefulness of paradata to analyze the quality of survey responses (Yan and Olson 2013). Compared with other survey modes such as face-to-face and mail surveys, web surveys provide an easy way to collect paradata. There is a clear need for research that makes use of paradata to investigate sources of measurement errors with respect to survey-based experiments in general and stated choice experiments in particular.

References

Buskirk, Trent D., and Charles Andrus (2012). Smart surveys for smartphone: exploringvarious approaches for conducting online mobile surveys via smartphones. *Survey Practice* (http://surveypractice.wordpress.com/2012/02/21/smart-surveys-for-smart-phones/).

Callegaro, Mario (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice* (http://surveypractice.wordpress.com/2010/12/08/device-respondent-has-used/).

Campbell, Danny, Morten Raun Mørkbak, and Soren Boye Olsen (2012). Response latency in stated choice experiments: Impact on preference, variance and processing heterogeneity. *Paper presented at the 19th Annual Conference of the European Association of Environmental and Resource Economists, Prague, 27 - 30 June 2012*.

Campbell, Danny, Morten Raun Mørkbak, and Soren Boye Olsen (2013). How quick can you click? Accommodating 'quick' responses to online stated choice questions. *Paper presented at the 15th Annual BIOECON Conference, 18-20 September 2013, Kings College, Cambridge United Kingdom*.

Dellaert, Benedict G.C., Bas Donkers, and Arthur van Soest (2012). Complexity Effects in Choice Experiment-Based Models. *Journal of Marketing Research* 49(3): 424-434.

DeShazo, J., and German Fermo (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123-143.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method, 3rd edition*. John Wiley: Hoboken, NJ.

Kontoleon, Andreas, and Mitsuyasu Yabe (2003). Assessing the Impacts of Alternative 'Opt-out' Formats in Choice Experiment Studies. *Journal of Agricultural Policy and Resources* 5: 1-43.

Kwak, Nojin, and Barry Radler (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics* 18(2): 257-274.

Lindhjem, Henrik, and Ståle Navrud (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics* 70(9): 1628-1637.

Lyberg, Lars E., and Paul P. Biemer (2007). Quality Assurance and Quality Control in Surveys, pp. 421-441, in: Edith D.de Leeuw, Joop J. Hox, and Don A.Dillman (Eds.), *International Handbook of Survey Methodology*. Mahwah, NJ: Lawrence Erlbaum Associates / Psychology Press.

Manfreda, Katja L., and VasjaVehovar (2007). Internet Surveys, pp. 264-283, in: Edith D.de Leeuw, Joop J. Hox, and Don A. Dillman (Eds.), *International Handbook of Survey Methodology* . Mahwah, NJ: Lawrence Erlbaum Associates / Psychology Press.

Marta-Pedroso, Cristina, Helena Freitas, and Tiago Domingos (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics* 62(3): 388-398.

Millar, Morgan M., and Don A. Dillman (2012). Encouraging Survey Response via Smartphones: Effects on Respondents' Use of Mobile Devices and Survey Response Rates. *Survey Practice* 5 (No. 4).

Morgan, Stephen L., and Christopher Winship (2007). *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Morrison, Mark, MacDonald, Darla H., Boyle, Kevin, Rose, John, and Roderick Duncan (2013). Investigating Differences between Internet and Mail Implementation of a Stated-Preference Study While Controlling for Differences in Sample Frames and Self-Selection Effects. *Paper presented at the International Choice Modelling Conference 2013*.

Nielsen, Jytte S. (2011). Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. *Resource and Energy Economics* 33(1): 119-129.

Olsen, Soren Boye (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics* 44(4): 591-610.

Peytchev, Andy, and Craig A. Hill (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review* 28: 319-335.

Poe, Gregory L., Kelly L. Giraud, and John B. Loomis (2005). Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87: 353-365.

Rosenbaum, Paul R., and Donald B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.

Scarpa, Riccardo, Thiene, Mara, and Kenneth Train (2008). Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics* 90: 94-1010.

Schaeffer, Nora C., and Stanley Presser (2003).The Science of Asking Questions.*Annual Review of Sociology* 29: 65-88.

Swait, Joffre, and Wiktor Adamowicz (2001). Choice environment, market complexity and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes* 86: 141-167.

Train, Kenneth (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press

Train, Kenneth E., and Melvyn Weeks (2005). Discrete choice models in preference space and willingness-to-pay space, pp. 1-16, in: Riccardo Scarpa and Anna Alberini (Eds.), *Application of simulation methods in environmental and resource economics.* Dordrecht: Springer.

Windle, Jill, and John Rolfe (2011). Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis and Policy* 41(1): 83-97.

Yan, Ting, and Kristen Olson (2013). Analyzing Paradata to Investigate Measurement Error, pp. 73–95, in: Frauke Kreuter, *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, NJ: John Wiley & Sons.